MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIAUNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2016/2017

HPB2031 – Bioinformatics Algorithms I

(All sections / Groups)

13 OCT 2016
2:30-4:30 PM
[2 hours]

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 3 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Please write all your answers in the Answer Booklet provided.
4. You may use calculator.

Question 1

a) What is Biological databases? Briefly explain the types of biological databases available and give an example for each type?

[4 marks]

b) What are the two popular dynamic programming algorithms used in sequence alignment? Briefly explain the three essential components of Needleman- Wunsch algorithm.

[4 marks]

c) Draw the F- matrix with example?

[2 marks]

Question 2

a) Calculate a BLOSUM matrix for the following sample:

Given a block of 4 A's, 1 B and 1 C

...A...
...A...
...B...
...A...
...C...
...A...

List all the amino acids pairs (substitutions - F_{xy}) and calculate the probability of each (P_{xy}). Calculate the expected frequency of A, B, C (q_x). Calculate the odds scores as observed probabilities / expected probabilities. Convert the score to log-odd score using base 2 with a scaling constant of 2. Assume no evolutionary history is known.

[6 marks]

b) Briefly discuss the steps involve in BLAST algorithm?

[4 marks]

Question 3

a) Given a set of four instances x1=(4,2), x2=(4,3), x3=(6,5) and x4=(8,8) , and given two initial clusters with centroids M1 =(3,3) and M2=(6,3) , by using the distance function $dist(x_i, x_j) = \sum_e |x_{i,e} - x_{j,e}|$ assign the instances into clusters and update the centroids in one iteration.

[5 marks]

Continued...

b) Briefly discuss the different types of methods involve in multiple sequence alignment with example.

[3 marks]

c) What is the difference between sequence homology and sequence similarity in sequence analysis?

[2 marks]

Question 4

a) Given the following distance matrix, construct the corresponding phylogenetic tree using the distance method (UPGMA). [7 marks]

	A	B	C	D	E	F
A	0					
B	5	0				
C	4	7	0			
D	7	10	7	0		
E	6	9	6	5	0	
F	8	11	8	9	8	0

b) What is a phylogenetic profile? [1 mark]

c) What inferences about the proteins can you make from the profile given below? [2 marks]

	Genome		
Protein	G1	G2	G3
P1	0	1	1
P2	0	0	1
P3	1	0	0
P4	0	1	1

Question 5

a) In the context of Pathways, what is a PGDB? What are the components (tiers) that exist in PGDBs? [4 marks]

b) Give two main differences between "pathway tools" and KEGG. [2 marks]

c) List all kinds of biological pathways or networks and state how they are related to "systems biology". [4 marks]

END OF PAPER